MEAN - COVARIANCE STRUCTURE MODELS IN ECONOMICS RESEARCH: AN APPLICATION TO A LENDING PROGRAM FOR DEVELOPMENT IN BURKINA FASO

PAXTON, Julia*
and THRAEN, Cameron

Abstract
As applied development economic models become more sophisticated, they include increasingly complex conceptual variables. Due to data collection limitations, accurate proxies and continuous variables are often unavailable. A Mean and Covariance Structure model (MECOSA) is offered as a useful methodology for the incorporation of latent variables with metric, censored metric, dichotomous and ordinal indicators. As an example, conceptual variables (including borrower homogeneity and the domino effect) presented in the Besley and Coate (1995) group lending repayment game were specified as latent variables with non-metric indicators. Data from 140 groups from a group lending program in Burkina Faso were used to demonstrate the application and interpretation of MECOSA.

JEL Classification: C5, C8, O16
Key words: mean and covariance structure models, latent variables, development finance, group lending

1. Introduction

As applied development economic models become increasingly sophisticated, more non-traditional economic and social variables have been incorporated into theoretical and empirical economic models. The economist's traditional approach to structural equation

* Julia Paxton Department of Economics, Ohio University, e-mail paxton@ohio.edu, and Cameron Thraen, Department of Agricultural, Environmental, and Applied Economics, The Ohio State University, USA.

modeling relies on the specification of the relevant economic model wherein all variables are observed constructs, such as income, price, or quantity, and these constructs are taken to be measured without error. Should a variable not be observable, it is replaced by an observable surrogate or "proxy" variable which is believed to be correlated closely with the unobserved variable and the estimation of the model parameters is accomplished using traditional methods. However, as Krasker and Pratt (1986) demonstrate, casual use of proxy variables can lead to serious parameter biases including incorrect parameter signs on the proxy variable(s), bias in the other included parameters and overall incorrect statistical inference.

In order to estimate empirical models more precisely, increasingly complex econometric tools are needed that can be flexible enough to include non-traditional variables and variables for which a non-metric scale is the only empirical counterpart. Increasingly, applied economists are turning to latent variable models that use multiple indicators of a variable that has no perfect proxy [Bauwens, L. and Veredas, D. (2004); Chesher (2003); Heckman, J. and Vytlacil, E. (2000)].

In the field of development finance, the study of loan repayment in solidarity groups[1] is particularly amenable to the incorporation of non-traditional economic variables since the repayment performance is directly related to the dynamics within the group. Group lending game theory [see Besley and Coate (1993)] has incorporated a variety of variables representing complex concepts that have no direct empirical measure such as (i) *group member homogeneity*, and (ii) the *domino effect* (one individual deciding to default given that the other defaults).

An appropriate statistical methodology must be able to handle unobserved and non-metric variables jointly, a task for which traditional single or system of equations econometric methods are not well suited. This purpose of this paper is (i) to propose the mean-covariance model as a useful statistical framework for applied economic models that incorporate complex variables and relationships, and (ii) to provide an application of the technique to a group loan repayment model using data from Burkina Faso. The statistical framework allows for the direct incorporation of conceptual variables and non-metric threshold variables into the analysis.

2. Modeling Methodology: Mean-covariance Structural Model

Applied economists are keenly aware of the many data definition and measurement problems inherent in the specification of meaningful economic models. These obstacles include modeling unobserved (or latent) endogenous variables, and the inclusion of threshold models for non-metric endogenous variables.[2] Non-metric threshold models include the familiar (i) metrically scaled or metrically classified, (ii) one or two sided censoring of endogenous variables (tobit type), and (iii) ordered categorical endogenous variables. Comprehensive economic models often times will necessitate a specification that incorporates many of these variable types in a simultaneous equations system.

Structural Equation Models

The structural equation modeling methodology with latent variables is widely known as the LISREL (LInear Structural RELations) model after Jöreskog (1977) although numerous software packages using structural equation models are available[3]. The LISREL method is an extension of the factor analysis model to incorporate linear structural relations among factors. Briefly stated the factor analysis model is of the form[4]

$$x = m + \Lambda z + u \quad \text{with } E(z)=0,\ V(z)=\Phi,\ E(u)=0,\ V(z)=\Theta,\ E(uz') \quad (1)$$

yielding the (unconditional) mean and covariance structure for x

$$E(x) = m,\ V(x) = \Lambda\Phi\Lambda' + \Theta \quad (2)$$

where the p x 1 vector variate x represents observed variables, the m x 1 vector variate z represents common factors, and the p x 1 vector variate u represents unique variables. The common factors, z and unique variables, u, are examples of latent variables. They cannot be observed but their existence is hypothesized to explain relationships between latent variables. The LISREL model extends this factor model by incorporating linear structural relations among factors (Cziraky, 2004). The vector of common factors z is partitioned into two subvectors, z_y and z_x, that satisfy the linear structural relations

$z_y = \mathrm{B}_y z_y + \Gamma z_x + e$, where $E(z_x) = 0$, $E(e) = 0$ and $Cov(z_x, e') = 0$.

Elements of z_y depend on z_x through Γ and the endogenous variables z_y and B_y. The observable variables y and the observable variables x are linked to the latent endogenous and exogenous factors by a *measurement* model

$$\begin{aligned} y &= m_y + \Lambda_y z_y + u_y \\ x &= m_x + \Lambda_x z_x + u_x \end{aligned} \qquad (3)$$

The LISREL model generates the covariance structure

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix} \qquad (5)$$

where the dependence of the covariance structure on the fundamental model parameters can be seen from the elements of Σ

$$\Sigma_{yy} = \Lambda_y (\mathrm{I} - \mathrm{B}_y)^{-1} (\Gamma \Phi_x \Gamma' + \Psi)(\mathrm{I} - \mathrm{B}'_y)^{-1} \Lambda'_y + \Theta_{uy}$$
$$\Sigma_{yx} = \Lambda_x (\mathrm{I} - \mathrm{B}_y)^{-1} \Gamma \Phi_x \Lambda'_y = \Sigma'_{xy}$$
$$\Sigma_{xx} = \Lambda_x \Phi_x \Lambda'_x + \Theta_{ux}$$

While structural equation modeling provides a mechanism to accommodate latent variables, it has some limitations in its flexibility[5]. Economic modeling requires a methodology that can incorporate not only unobserved variables but also the mapping of unobserved variables onto observed non-metrically scaled variables. Mean-covariance structural analysis (MECOSA) is one such method that can incorporate these characteristics.[6] By the use of threshold models the MECOSA method can directly incorporate non-metric, i.e. binary, categorical, ordered categorical, etc., and censored dependent variables into the structural modeling. Unlike LISREL that incorporates unconditional mean and covariance structures, MECOSA uses conditional structures.

A mean and covariance structure can be expressed as being composed of three parts. The first part will be familiar as the structural model defining the relationship between the relevant endogenous latent variable vector η_i and latent exogenous variables x_i:

$$h_i = \mathrm{B} h_i + \Gamma x_i + V \qquad (9)$$

where B is the matrix of regression coefficients for the endogenous variables, Γ is the matrix of regression coefficients for the explanatory variables, and ς_I is a vector of disturbances with expected value 0 and covariance matrix ψ.

The second part is the measurement model that links observable variables with their unobserved counterparts. The measurement model is essentially a joint factor analytic model specified as:

$$y_i = n + \Lambda h_i + d \qquad (10)$$

where y_i and x_i are observed (indicator) variables. Λ is the matrix of factor loadings which specifies the link between the observed concept variables h and the observed variables y. The vector d represents the measurement errors for y_i, with $\mathrm{E}(d_i) = 0, V(d_i) = 0, E(d_i, h_i) = 0.$ The parameter vector of interest contains the unrestricted elements in $\mathrm{B}, \Gamma, \Psi, u, \Lambda$ and Θ.

In the MECOSA structure, it is the measurement model which explicitly provides the link between unobservable concept variables in the structure and the observed or indicator variables in the field. The reduced form parameters are given by

$$g(J) = u$$

$$\Pi(J) = \Lambda(\mathrm{I} - \mathrm{B})^{-1}\Gamma$$

$$\Sigma(J) = \Lambda(\mathrm{I} - \mathrm{B})^{-1}\Psi\Lambda(\mathrm{I} - \mathrm{B})^{\prime -1}\Lambda' + \Theta$$

The third feature of this structure is that with the MECOSA approach a wide class of specifications on the dependent variables can be accommodated. A dependent variable may be (i) metrically scaled and as such is identical to the unobserved endogenous variable; (ii) metrically classified and thus represent ordinal groupings; (iii) one or double sided censored and so encompass the familiar Tobit types; and (iv) ordered categorical with unknown threshold values and thus allow for ordinal probit type relationships. The applied economist will recognize the value of this approach as it encompasses many of the variable types encountered in empirical analysis.

Each element of the observed endogenous vector is linked to an underlying latent endogenous variable through *threshold models*. Each element $y_{j,i}$, $i = 1,\dots,n$ of the observed endogenous vector $\boldsymbol{y_j}$ is linked to an element y^*_{ji} of the unobserved or latent variable $\boldsymbol{y^*_j}$ by the appropriate threshold model (see below).[7] The vector of threshold values of an endogenous variable is given by $\boldsymbol{t}_i = (\boldsymbol{t}_{i,1}, \boldsymbol{t}_{i,2}, \dots, \boldsymbol{t}_{i,K_i+1})'$. The endogenous variable thresholds are denoted in the vector $\boldsymbol{t(J)} = (t'_1, \dots, t'_n)'$ and are a function of the structural parameter vector J. (Arminger, Wittenberg, and Schepers (1996))

Consider the following threshold models that are frequently encountered in applied economic research:

metrically scaled endogenous variable y_{ji}: $y_{ji} = y^*_{ji}$ (14)

metrically classified endogenous variable with known class boundaries: $t_{i,1} < t_{i,2} < \ldots . < t_{i,K_i}$ and $K_i + 1$ categories (15)

$$y_{ji} = k \Leftrightarrow y^*_{ji} \in [t_{i,k-1,} t_{i,k})$$

$$[t_{i,0}, t_{i,1}) = (-\infty, t_{i,1}) \text{ and } t_{i,K+1} = +\infty.$$

one sided censored endogenous variable where the threshold of the variable is known *a priori*:

$$y_{ji} = \begin{Bmatrix} y^*_{ji} \text{ if } y^*_{ji} > t_{i,1} \\ t_{i,1} \text{ if } y^*_{ji} > t_{i,1} \end{Bmatrix} \tag{16}$$

ordered categorical endogenous variable with unknown threshold values $t_{i,1} < t_{i,2} < \ldots . < t_{i,K_i}$ and K_i+1 ordered categories (ordered Probit) (17)

$$y_i = k \Leftrightarrow y^*_i \in [t_{i,k-1,} t_{i,k})$$

$$[t_{i,0}, t_{i,1}) = (-\infty, t_{i,1}) \text{ and } t_{i,K+1} = +\infty.$$

The MECOSA methodology can easily be recognized as a generalization of the traditional simultaneous equation structure familiar to economists. When Λ_y and Λ_x are identity matrices, and the covariances of the measurement errors, Θ_{dy} and Θ_{dx}, are identically zero, the above model reduces to the traditional econometric specification. If B is not null, then the model represents the traditional simultaneous equations system. The value of the MECOSA modeling structure is that it can accommodate the diverse nature of the variables without resorting to unrealistic assumptions as to absence of measurement error and/or the specification of proxy variables. This approach encompasses traditional simultaneous econometric models and factor analytic models and can represent a much broader range of model structures likely to be encountered in applied economic research. The MECOSA estimation strategy is based on the assumed multivariate normality of y_i^* given x_i:

$$P(y_i^* | x_i) = (y_i^* | g\,(J) + \Pi(J) x_i,\ \Sigma(J)) \tag{18}$$

is the conditional density of y_i^* given x_i with expected value $g\,(J) + \Pi(J)\ x_t$ and covariance matrix $\Sigma(J)$. Estimation of the parameters can be accomplished by application of a two stage process using first a limited information ML estimator and then a minimum distance estimator. Statistical estimation of the mean-covariance model requires the simultaneous estimation of the parameters of the structural model, the measurement model and the covariance structure of the disturbance terms. In recent research, MECOSA has been used to specify models for finite mixtures of multivariate normal densities [Arminger, Stein, and Wittenberg, (1999)]. The MECOSA statistical estimation methodology is based on the work of Muthén (1979, 1983, 1984) and generalized by Küsters (1987).[8]

3. An Application of MECOSA in a Model of Group Lending Repayment

In order to demonstrate the effectiveness of the MECOSA approach for applied economists, a simple empirical model of group lending is presented. In the model, both observed and latent variables with non-metric indicators are used.

3.1 A Model of Group Lending Repayment

Besley and Coate (1995) developed a group loan repayment game based on the widely popular group lending programs in developing countries. In the game, two homogeneous borrowers have loans with joint liability and receive a random return on their project. In one possible game outcome, one individual independently decides to repay his/her part while the other does not. Once the repayment problem has been exposed to both members, the correct paying borrower may repay both portions of the loan or may decide to default on his/her part as well (*domino effect*).

In order to empirically test some of the concepts of group dynamics presented in the Besley and Coate repayment game, a two stage estimation can be constructed. Since only some group members may have repayment difficulties, the first stage examines the determinants of having a repayment problem. A repayment problem may have arisen even in groups with perfect repayment records. For example, certain groups may have had one or several group members with repayment problems but the loan was repaid on time and the microfinance institution never became aware of the problem. Once a problem has been identified, a second stage analyzes the factors leading to successful repayment of the loan. The independent variables in the model will depend on the context and degree of model complexity. These two stages were specified and tested using data from the group lending institution Projet de Promotion du Petit Credit Rural (PPPCR) in Burkina Faso. The first stage of the group repayment model[9] is ideally suited to present the MECOSA methodology since it has a relatively simple structure, yet incorporates observed, unobserved (latent), metric, and non-metric variables in a simultaneous equation system with the following specification:

Structural Equations

$$h_0 = b_{01}h_1 + b_{02}h_2 + g_{00}x_0 + g_{01}x_1 + z_0 \qquad (19)$$

$$h_1 = b_{10}h_0 + g_{12}x_2 + z_1 \qquad (20)$$

Measurement Model

$$Y_0 = n_0 + l_0 h_0 + d_0 \qquad (21)$$

$$Y_1 = n_1 + l_1 h_1 + d_1 \qquad (22)$$

$$Y_2 = n_2 + l_2 h_1 + d_2 \qquad (23)$$

$$Y_3 = n_3 + l_3 h_2 + d_3 \qquad (24)$$

$$Y_4 = n_4 + l_4 h_2 + d_4 \qquad (25)$$

Endogenous variables

h_0 = problem (observed, non-metric) h_1 = domino effect (unobserved) h_2 = homogeneity1 (unobserved)

Exogenous variables

x_0 = urban (observed, binary) x_1 = other credit1 (observed, metric) x_2 = loan cycle1 (observed, metric)

Latent variable observed indicators

$Y_0 =$ problem (observed, non-metric) $Y_1 =$ sector domino (observed, non-metric) $Y_2 =$ group domino (observed, metric) $Y_3 =$ homogeneity family(observed, metric) $Y_4 =$ homogeneity scale (observed, metric)

3.2 Data generation and description

A questionnaire was designed to probe into the group dynamics of 140 women's solidarity groups that borrow from the PPPCR in Burkina Faso. In addition to concepts presented in the Besley and Coate repayment game (such as borrower homogeneity and the domino effect), context specific variables were incorporated. As in many field surveys in applied economic research, a combination of metric and non-metric variables was gathered. In addition, multiple indicators of latent variables were collected rather than relying on a single index or proxy assumed to be a perfect indicator of the latent variable.

The use of latent variables

Two of the variables discussed in the Besley and Coate repayment game were member homogeneity and the domino effect. Measuring these variables empirically is difficult since no one proxy represents these concepts. Therefore, two observable indicators of each of these latent variables were used in order to incorporate measurement error and avoid model misspecification and the resulting inconsistency and bias associated with a single proxy assumed to have no measurement error. The domino effect refers to a chain reaction of default in group lending encouraged by the joint liability that prevents anyone in the group to receive future loans if even one person is delinquent. In the case of the PPPCR, two types of domino effect are present and therefore, two different indicators are used. One indicator variable is a binary 0/1 variable that reports whether or not any group in the sector has experienced default. In the structure of the PPPCR, if one group in the sector defaults, then all of the groups are barred from receiving future loans. This would provide an incentive for consistent payers to default because the perceived future value of this line of credit would be zero. The second indicator variable is the number of individuals in the group with repayment problems.

This is a metric variable ranging from 0 to 5 members. Another latent variable that potentially could affect the occurrence of a repayment problem is the *homogeneity* of the group. Since *homogeneity* is a pure concept that is difficult to measure using only a proxy variable assumed to have no measurement error, indicator variables of homogeneity were specified. One indicator was a *homogeneity index* constructed out of the 10 yes/no questions asking if all of the members of the group were of the same ethnic group, had the same occupation, had similar incomes, lived in the same neighborhoods, etc. in the survey.

The other indicator of *homogeneity* was a metric variable measuring the number of different families that comprised the group of 5 women. In rural Burkina Faso, where large families and polygamy exist, it was common to have several immediate family members or co-wives in the same group.

The use of metric and non-metric variables

If metric indicators were available for all latent variables and dependent variables were also metric, a LISREL model could be used. However, the presence of both metric and non-metric dependent variables and indicators requires the use of a MECOSA methodology that incorporates threshold models to link the non-metric variables to an underlying continuous function. The dependent variable *problem* was measured as a binary (yes/no) variable. In the questionnaire, group members were asked if during the course of their current loan if one or more members had ever experienced some problem related to their work or personal lives that resulted in difficulty repaying the loan. The MECOSA methodology allows for the use of a binary dependent variable similar to a Probit model, but in addition allows for the incorporation of latent variables. One of the indicators of the latent variable *domino effect* was also a non-metric, binary variable. The presence of default anywhere in the sector was one indicator of *domino effect*. As such, both the *domino effect* and *problem* variables were specified as ordered-categorical variables (with two categories) and were linked to an underlying continuous function with the ordered Probit threshold model specified in equation (17).

3.3 Description of the structural equations and measurement model

The structural equations define the series of relationships between the dependent and independent variables through a set of simultaneous equations. The primary equation is the first equation (19) that examines the determinants of an individual in a group having a repayment problem. *Problem* is dependent on observed exogenous variables (an *urban* rural dummy variable and the number of *other credit*s outstanding the group member has), an exogenous latent variable (*group member homogeneity)*, and an endogenous latent variable (*domino effect)*[10].

Domino effect (equation 20) is determined simultaneously with *problem* in this model. This simultaneity comes about because of the dynamics inherent in the credit program. If other members of the group or the sector have already defaulted, other members may choose to shirk. In addition, the *domino effect* is a function of the exogenous variable, *loan cycle*. In studying group lending and repayment it is apparent that as the loan cycles progress, groups tend to default more frequently. This in turn can lead to the presence of the *domino effect*. The loan cycle variable ranged from 1 to 6 loan cycles.

The measurement model equations (21-25) are inserted into the structural equations. In the case of an endogenous variable with a single indicator (such as *problem*), the measurement model provides no additional information. Thus, in equation (22), the constant is equal to zero, the coefficient lambda is unitary thereby equating the latent variable Y with the endogenous variable eta, and the error term is assumed to be zero. In other words, these are not latent variables, but simple endogenous variables with a single proxy. However, in the case of the true latent variables, no such simplification occurs. *Homogeneity* and the *domino effect* are unobserved variables in the model and as such, require the specification of indicators (equations 22 through 25) for the measurement model.[11] The specification of two or more indicators allows for the incorporation of measurement error.

3.4 Analysis of estimation results

The empirical interpretation of a mean and covariance structure model can be divided into the covariance structure model specified by the measurement model and the mean structure model specified by the structural equations. The parameters of primary interest to the overall interpretation of the group lending model are the in the structural equations since they frame the entire analysis. However, it is useful to begin by interpreting the measurement model to facilitate the analysis of the latent variables in the structural equations. Equations 26 through 29 present the coefficients and standard errors of the measurement model using the variable names rather than symbols for clarity.

$$E(\text{Sectoral domino effect}) = 1.0 \text{ domino effect} \quad (26)$$

$$E(\text{Individual domino effect}) = \underset{(0.1629)}{2.2699} + \underset{(0.1509)}{1.8679} \text{ domino effect} \quad (27)$$

$$E(\text{Family Homogeneity}) = \underset{(0.0867)}{0.5332} + \underset{(0.1488)}{0.3574} \text{ homogeneity} \quad (28)$$

$$E(\text{Homogeneity Scale}) = 1.0 \text{ homogeneity} \quad (29)$$

The number of members within a group that experienced repayment problems was one indicator of the domino effect. Whether or not any other group in the sector had arrears was a second indicator of the domino effect. By setting the constant equal to zero and the coefficient equal to one in the sectoral domino effect equation, it is possible to scale both domino effect indicators so that they may be compared and the model is identified. With these restrictions, the individual domino effect has the same expected value as the sectoral domino effect. Also, a one unit change in the unobserved variable domino effect has a one unit expected change in each of the indicators. The unobserved variable domino effect has the same scale[12] as sectoral domino effect. The domino effect parameter estimate from the individual domino effect equation is 1.8679, which indicates that a one unit change in the unobserved variable domino effect will increase the expected value of the individual domino effect by 1.8679. A similar interpretation can be given to the unobserved variable homogeneity. Its two indicators include the number of different families in the group and a scale of 10 questions relating to member wealth, family, age, etc.

The homogeneity scale was set as the basis for comparison, with a zero intercept and coefficient of one. As homogeneity changes by one unit, the expected value of family homogeneity changes by 0.3574. Each of the parameters estimated in the measurement model were highly significant at the a =0.01 level. For empirical research, the structural parameter estimates provide the primary source of insight into the relationship between the dependent and independent variables.

In this case, the analysis centers around the equation defining the potential factors that determine repayment problems. Because the dependent variable *problem* is binary, the mean structure model is a structural probit model and its parameters can be interpreted equivalently to a probit model.[13] The estimated model parameters and the marginal effects of the *problem* equation are given in Table 1.

The parameter estimates given in Table 1 are measures of the derivatives of the conditional mean function. The parameter estimates provide an approximation to the change in probability as the dependent variable equals one at the regressor means. It is useful to calculate the marginal effects of a probit model since they can be utilized to show how a unit change in one of the variables will affect the probability of the dependent variable equaling one. For example, a one unit change in the variables *homogeneity* and *urban* have the largest effects on the probability that a problem will arise.

The elasticities are measured at the mean values of the independent variables. High elasticity indicates that a slight variation in the variable's value will result in a significant change in the probability of the dependent variable equaling one. The *domino effect* is the most elastic variable measured and thus, if an objective of lending program managers is to reduce the number of problems that their groups experience, they should examine ways to reduce this *domino effect*.

Several important findings result from the estimation that examines the determinants of repayment *problems* arising in groups. Both the endogenous latent variable (*domino effect)* and the exogenous latent variable (*homogeneity)* were found to be significant. The urban exogenous dummy variable was highly significant while the variable *other credit* was not significant.

Not surprisingly, the urban dummy variable was negative with a low standard error and strongly significant. Problems tended to occur more in rural areas. The rural clients have a higher dependence on agricultural activity as the base of the rural economy. In conjunction with the reliance on agriculture comes a higher degree of income variation, risk, and covariant incomes. In contrast, urban markets tend to be more diversified with a greater degree of monetization.

The *homogeneity* coefficient is positive indicating that the more homogeneous the group, the more problems occurred. One likely explanation for this phenomenon is the existence of co-variant risk in homogeneous groups with similar economic activities and social relations. In addition, due to the domino effect, if one member of a tight-knit group sees the others experiencing problems, she also may shirk her responsibilities, leading to problems for herself.

The *domino effect* had a significant positive influence on having problems as expected. As other groups and members defaulted, more problems arose. Part of this effect could be due to shirking as members decide that the marginal benefits of repaying exceed the marginal costs of repaying. In the second simultaneous equation (20), the domino effect was dependent on *problem* and the *loan cycle*. The coefficients on both variables were found to be positive as expected (0.6295 for problem and 0.1018 for loan cycle) and significant at the $a = 0.05$ level.

3.5 Comparing MECOSA and a Probit model without latent variables

In this example, the MECOSA model included a latent endogenous variable (the domino effect), a latent exogenous variable (group homogeneity), and two exogenous variables. Table 2 shows the estimation results for a probit model assuming all exogenous variables to have no measurement error. Rather than having multiple indicators for the latent variables, single proxies were selected to represent the concepts. "Sectoral" domino effect was selected to represent the domino effect and the number of different families in the group was chosen to represent group homogeneity. While the signs of the coefficients remain the same when comparing the MECOSA and probit estimations, the homogeneity variable is not significant in the simplified version, thereby skewing the interpretation of the results.

4. Conclusion

Mean- and Covariance structural models offer the potential to model the complex structures found in applied economics in a way that is more robust than traditional econometric methods. By using single proxies rather than multiple indicators of conceptual variables, researchers incorrectly are assuming measurement error to be zero and as a result, the results could be biased, inconsistent, and even give the wrong signed coefficient. In addition to being able to estimate a standard econometric specification, the MECOSA model can allow for the incorporation of latent variables with multiple indicators and threshold models for non-metric variables. In this respect, it is one of the most flexible methodologies available to applied economists. In order to illustrate the usefulness of the MECOSA methodology, the technique was applied to an empirical model of the determinants of repayment problems in group lending. The model included several latent variables with both metric and non-metric indicators. The incorporation of latent variables allowed for a greater depth of model specification and interpretation. Both latent variables in the model (*homogeneity* and the *domino effect*) were shown to be significant determinants of repayment problems. If a simple probit model is estimated omitting the multiple indicators, then homogeneity is no longer significant.

Despite the flexibility and sophistication of the MECOSA methodology, it has not received widespread application in economics and is not widely taught as part of the standard graduate econometric curriculum. The method is more than an elementary extension of the traditional mean structure methods such as classical regression with observed variables and as such requires much more by way of computer software and hardware for its implementation. Fortunately with contemporary computer technology, this is no longer a relevant constraint. For the practicing applied economist confronting important research issues formulated on conceptual variables and data with metric and non-metric measures in the field of applied economics, mean-and-covariance structural modeling provides a more comprehensive and robust econometric technique.

BIBILIOGRAPHY

Arminger, G., Clogg, C.C., and M.E. Sobel, eds., (1995) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press.

Arminger, G., Stein, P., and Wittenberg, J. (1999) Mixtures of Conditional Mean and Covariance-Structure Models. *Psychometrika* Vol. 64(1), pp. 475-494.

Arminger, G., Wittenberg, J. and A. Schepers, (1996) *MECOSA 3: A Program System of the Analysis of General Mean- and Covariance Structures, User Guide*, Friedrichsdorf, Germany: Aptech Systems, Inc.

Bauwens, L. and D. Veredas, (2004) The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations. *Journal of Econometrics*, Vol. 119 (2), pp. 381-412.

Besley, T. and S. Coate, (1995) Group Lending, Repayment Incentives and Social Collateral. *Journal of Development Economics*. Vol. 46(1), pp. 1-18.

Bollen, K.A., (1989) *Structural Equations with Latent Variables* New York: John Wiley & Sons.

Chesher, A. (2003) Identification in Nonseparable Models. *Econometrica*, Vol. 71 (5), pp. 1405-41

Cziraky, Dario, (2004) LISREL 8.54: A Program for Structural Equation Modelling with Latent Variables. *Journal of Applied Econometrics*, Vol. 19 (1), pp. 135-41.

Heckman, J. and Vytlacil, E., (2000) The Relationship between Treatment Parameters within a Latent Variable Framework. *Economics Letters*, Vol. 66(1), pp. 33-39.

Jöreskog, K.G., (1977) Structural Equation Models in the Social Sciences: Specification, Estimation, and Testing. in Krishnaiah, P.R., ed., *Applications of Statistics*, North-Holland, Amsterdam.

Muthén, B.O., (1979) A Structural Probit Model with Latent Variables. *Journal of the American Statistical Association*, Vol. 74, pp. 807-811.

Muthén, B.O, (1983) Latent Variable Structural Equation Modeling with Categorical Data. *Journal of Econometrics*, Vol. 22, pp. 43-65.

Muthén, B.O, (1984) A General Structural Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators. *Psychometrika*, Vol. 49, pp. 115-132.

Paxton, J., Graham, D., and Thraen, C., (2000) Modeling Group Loan Repayment Behavior: New Insights from Burkina Faso. *Economic Development and Cultural Change*, Vol. 48(3), pp. 639-55.

Wall, M. and Amemiya, Y., (2000) Estimation for Polynomial Structural Equation Models. *Journal of the American Statistical Association*, Vol. 95(451), pp. 929-40.

Table 1. Estimated Coefficients, Marginal Effects and Elasticities

Variable	Type of variable	Coefficient	Marginal Effect, Elasticity
Homogeneity	latent with 2 indicators	0.0822*	0.1798 0.3266
Domino Effect	latent with 2 indicators	0.8585*	0.0172 0.0311
Urban	exogenous binary	-0.6791*	-0.1422 -0.0920
Other Credit	exogenous metric	0.0410	0.0086 0.0070

elasticity at mean value * significant at $a = 0.05$ level

Table 2. Probit estimation with no latent variables

Variable	Type of variable	Coefficient (s.e.)
Homogeneity	exogenous ordinal scale	0.106 (0.079)
Domino Effect	exogenous binary	1.015* (0.183)
Urban	exogenous binary	-0.402* (0.175)
Other Credit	exogenous metric	0.160 (0.100)

* significant at $a = 0.05$ level

[1] Popularized by the Grameen Bank in Bangladesh that uses group lending to provide financial services to over 2,000,000 poor women, solidarity group lending uses small groups (often five members). Access to additional group loans is contingent on each member repaying his/her share and therefore mechanisms of peer pressure and helping behavior become important determinants of repayment.

[2] Bollen provides the following definition of latent random variables. "Latent random variables represent unidimensional concepts in their purest form. Other terms for these are unobserved or unmeasured variables and factors. The observed variables or indicators of a unobserved variable contain random or systematic measurement errors, but the unobserved variable is free of these. Since all unobserved variables correspond to concepts, they are hypothetical variables."

[3] For example: AMOS, CALIS, COSAN, EQS, LINCS, LISCOMP, Mx, PLS, RAM, RAMONA, SEPATH, STREAMS, TETRAD II

[4] This exposition follows closely that of Browne, M.W. and G. Arminger, (1995). 'Specification and Estimation of Mean- and Covariance-Structure Models', in Arminger, G., Clogg, C.C. and M.E. Sobel, eds., *Handbook of Statistical Modeling for the Social and Behavioral Sciences* Plenum Press, New York, pp. 185 - 249.

[5] Wall and Amemiya (2000) point out that the fitting procedures for structural equation models are only available for a limited class of models and offer a systemic statistical procedure to improve flexibility.

[6] MECOSA represents the general approach which is the focus of this paper. The term MECOSA also represents the MECOSA 3.0 software program authored by Gerhard Arminger, Jorg Wittenberg, and Andreas Schepers. In this paper the use of the term MECOSA should be understood by the reader to mean the general modeling method and not the specific software program. Two recently developed programs allow for greater flexibility in allowing for non-metric variables. Muthén has developed a user-friendly extension of LISREL (Mplus Version 2) that incorporates non-metric observed variables. MECOSA is a further generalization of the LISREL model in that it allows arbitrary models for conditional mean and covariance structures and arbitrary non-linear and linear restrictions while allowing for more flexibility in model specification.

[7] Note that in the first case the vector $\mathbf{y_i}$ is observed and is an indicator of the latent vector h. Now we are introducing the additional assumption that the indicator vector $\mathbf{y^*_i}$ is not observed.

[8] A complete treatment of the specification and estimation of mean-covariance models can be found in "Specification and estimation of mean- and covariance-structure models," Chapter 4 by Michael W. Browne and Gerhard Arminger in the *Handbook of statistical modeling for the social and behavioral sciences*, edited by Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel. New York : Plenum Press, 1995.

[9] For an estimation and discussion of both stages, see Paxton, J., Graham, D., and Thraen, C., "Modeling Group Loan Repayment Behavior: New Insights from Burkina Faso" Economic Development and Cultural Change, April, 2000.

[10]The MECOSA methodology treats all latent variables with indicators as endogenous for estimation purposes. (Arminger, G., Wittenberg, J., and Schepers, A. MECOSA 3: Mean and COVariance Structure Analysis User Guide, ADDITIVE GmbH, Friedrichsdorf, Germany, 1996.

[11] While an indicator variable can be thought of as a type of proxy for the unobserved variable, the departure here from the more traditional methodology is that this approach makes the link explicit by the measurement model and there are at least two indicators of the same unobserved concept to ensure identification of the model parameters (Bollen).

[12] Scaling of all unobserved variable indicators to a single indicator is useful for interpretation since each indicator may have a different measurement scale. The unobserved variable itself is put on the same scale as one of its indicators and then all other indicators can be compared to that reference scale.

[13] The probit model, based on the normal distribution can be written as:

$$PROB(Y = 1) = \int_{-\infty}^{b'x} f(t)dt = \Phi(b'x)$$

The coefficients b are not marginal effects typical in other estimations. In order to calculate the marginal effects, the standard normal density function is evaluated so that $\partial E[y]/\partial x = f(b'x)b$